

Mammalian
Gene
Collection



MGC: Its Status and Future

Interactome Networks

August 30, 2006

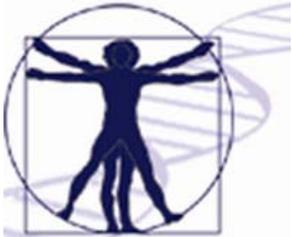
Daniela S. Gerhard, Ph.D.

Director, Office of Cancer Genomics

National Cancer Institute

Acknowledgements:

Mammalian
Gene
Collection



- Gary Temple
 - National Human Genome Research Institute
- Lukas Wagner, Wonhee Jang
 - National Center for Biotechnology Information
- Christa Prange Pennacchio
 - Lawrence Livermore National Laboratories, IMAGE
- MGC Project Team
 - 100+ individuals; for a complete list see:
 - Gerhard et al., Gen Research 14:2121, 2004



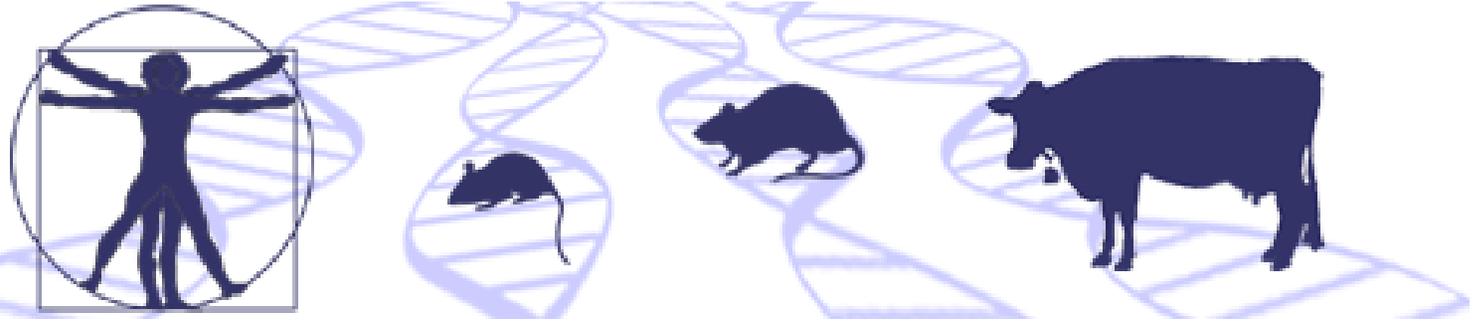
Outline

Mammalian
Gene
Collection



- FL clones from cDNA libraries (Phase I)
- PCR Generation of Human & Mouse Full-CDS Clones (Phase II)
- De novo Predictions of Human Exons, Genes, & Full-CDS Sequences (Phase III)
- Results of DNA Synthesis Pilot & Follow-up for Human and Mouse (Phase IV)
- Human ORFeome Collaboration

Mammalian
Gene
Collection



MGC: Phase I

XGC

ZGC

Method:

Mammalian
Gene
Collection



- Unidirectional cDNA libraries made by a number of academic laboratories and contractors
- 5' reads from 5,000 – 20,000 clones
 - Xenopus also has 3' reads
- 5' EST used to predict initiator methionine by 4 alternative methods
 - Match to RefSeq
 - Protein comparison
 - GenomeScan comparison
 - HKScan starts (human and mouse only)

Current Status

Mammalian
Gene
Collection



mgc.nci.nih.gov

21-Aug-06	Human	Mouse	Rat	Cow
Total MGC full ORF clones	24,014	19,096	5,058	4,680
Non-redundant genes	14,172	13,332	4,773	4,130
Candidate clones for full-length sequencing	3,054	1,361	794	4,076
Goal	18,368	18,471	6250	10,000
% of goal	77.2	72.2	76.4	41.3



IMAGE Consortium Distributors

[MGC Home](#)

Clone Info

- [Where to Buy](#)
- [Vectors & Method Overviews](#)

Sequencing Info

- [Candidate Clones for Genes](#)
- [MGC ESTs](#)

MGC Info

- [Project Summary](#)
- [Project Teams](#)
- [NIH Institutes](#)
- [References](#)

Other Species Collections

- [Danio \(ZGC\)](#)
- [Xenopus \(XGC\)](#)

➤ American Type Culture Collection

➤ Invitrogen, Inc

➤ Open Biosystems

➤ MRC Geneservice

➤ Resource Center of the German Human Genome Project

Current Status

xgc.nci.nih.gov

Under the auspices of NIHCD (S. Klein & D.S. Gerhard)
Collaboration with JGI: Erika Lindquist and P. Richardson

Mammalian
Gene
Collection



21-Aug-06	<i>laevis</i>	<i>tropicalis</i>
Total XGC full ORF clones	8,867	3,528
Non-redundant genes	8,422	2,909
Candidate clones for full-length sequencing	661	3,807
Goal (estimates)	9,200	9,000
% of goal	91.5	32.3



- Analysis: 8049 FL cDNA from *X. laevis*
2918 FL cDNA from *X. tropicalis*
- Results: 445 gene trios (allotetraploidization-derived *X. laevis* gene pair and their shared *X. tropicalis* ortholog)
- ➡ d_N/d_S comparisons within trios show evidence for purifying selection
 - ➡ d_N/d_S ratios within *X. laevis* are significantly elevated, i.e. relaxation of selective pressures on duplicated gene pairs

Current Status

Mammalian
Gene
Collection



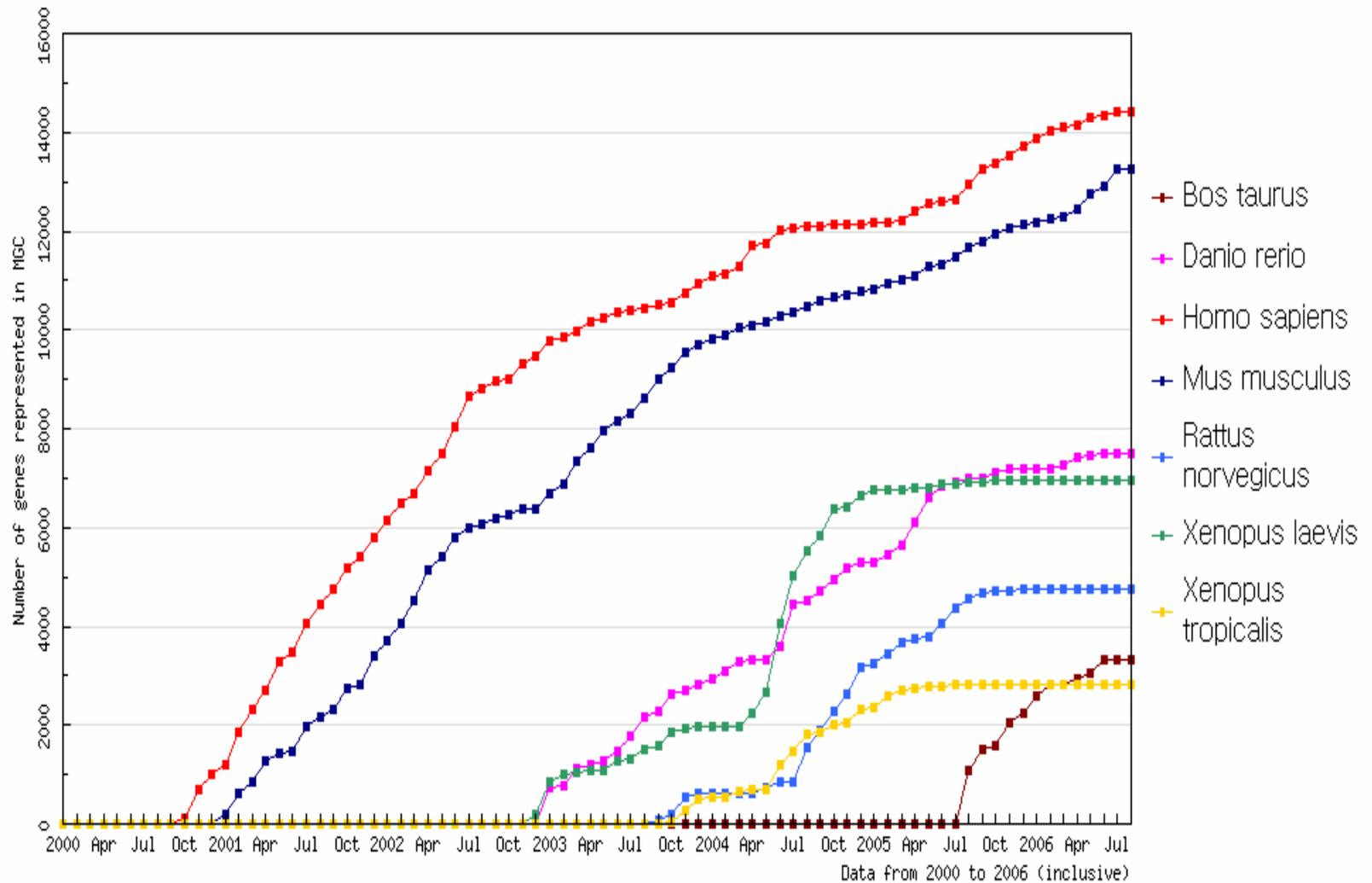
zgc.nci.nih.gov

Under the auspices of NIDDK R. Rasooli

21-Aug-06	#
Total ZGC full ORF clones	8,771
Non-redundant genes	7,640
Candidate clones for full-length sequencing	1,173
Goal	15,000
% goal	50.9

MGC Progress as of August 4, 2006

Mammalian
Gene
Collection



Mammalian
Gene
Collection



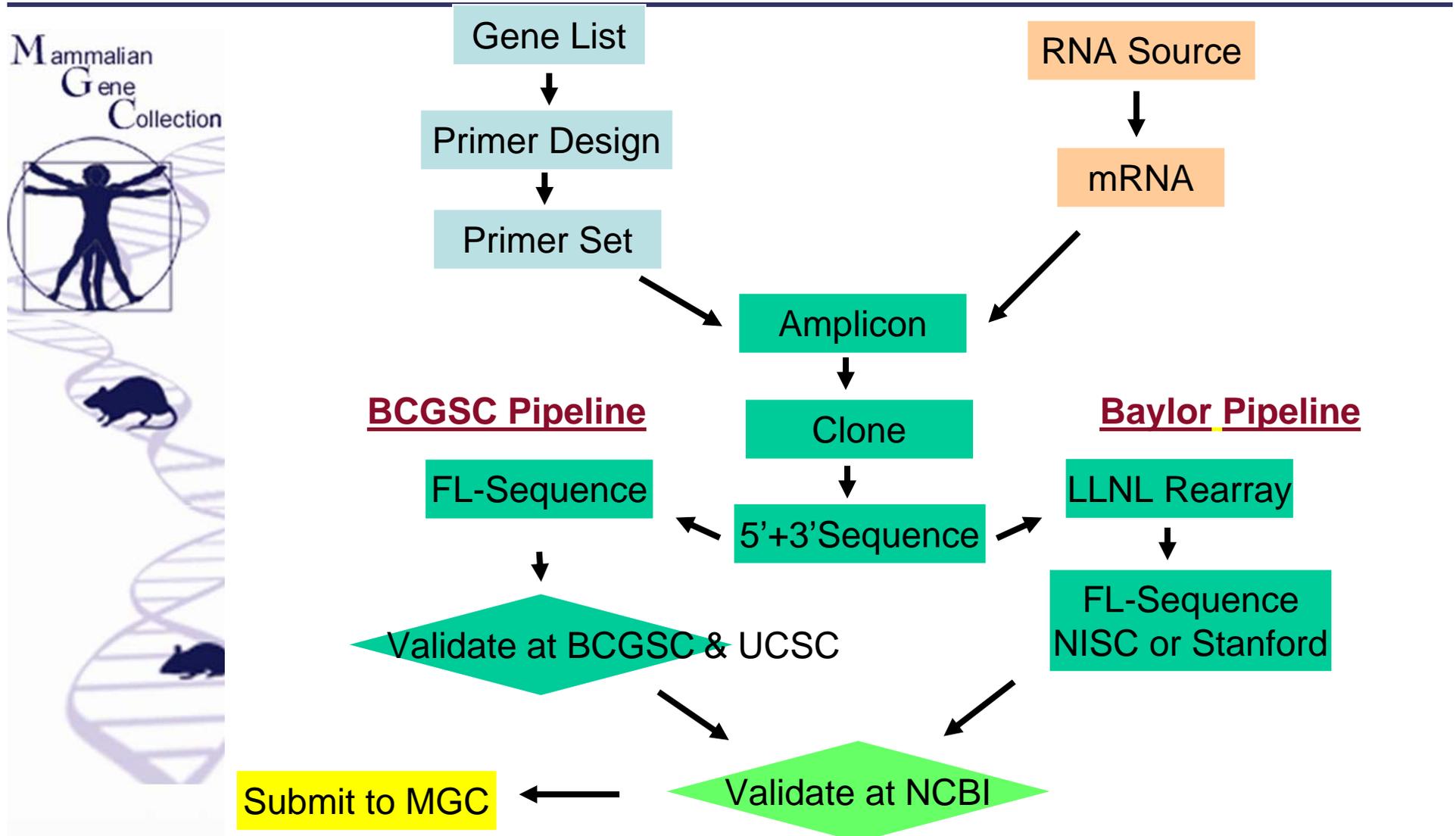
MGC: Phase II

PCR-based cloning

Pilot: 2003

Full scale: 2004

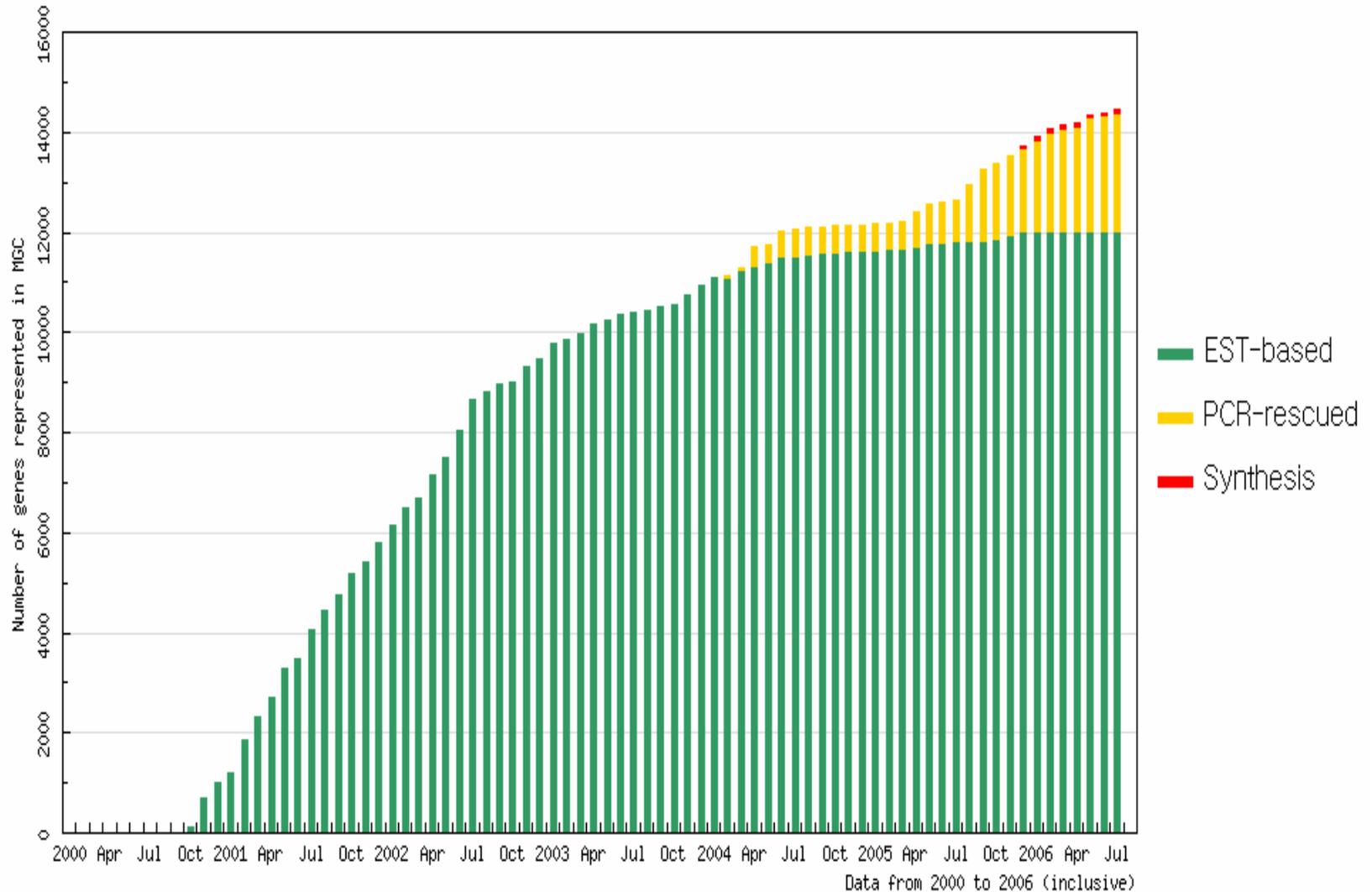
ORF Cloning by PCR Rescue



(Adapted from Hirst, M, 2005)

Progress of *Homo sapiens* clones by method

Mammalian
Gene
Collection



Accomplished

Mammalian
Gene
Collection



➤ For ~80% completed 2 attempts with a composite success rate between 50-60%

➤ # of genes targeted obtained*:

– Human	6,462	3,032
– Mouse	6,003	1,984
– Total	12,465	5,016

*as of August 25, 2006

Issues Encountered

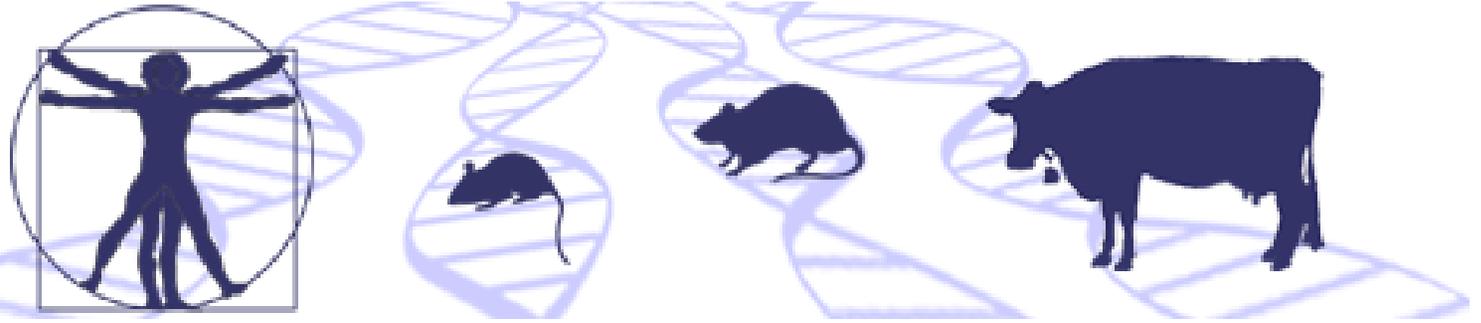
Mammalian
Gene
Collection



- Targets were not stable
 - XM_# vs. NM_#
 - Duplicate selections
- Alternative splicing
 - Effect on QC process
- Tracking for one group—3 sites
 - Cloning and EST generation
 - Picking of clones for FL
 - FL sequencing on 3rd
- Labor intensive in all steps



Mammalian
Gene
Collection

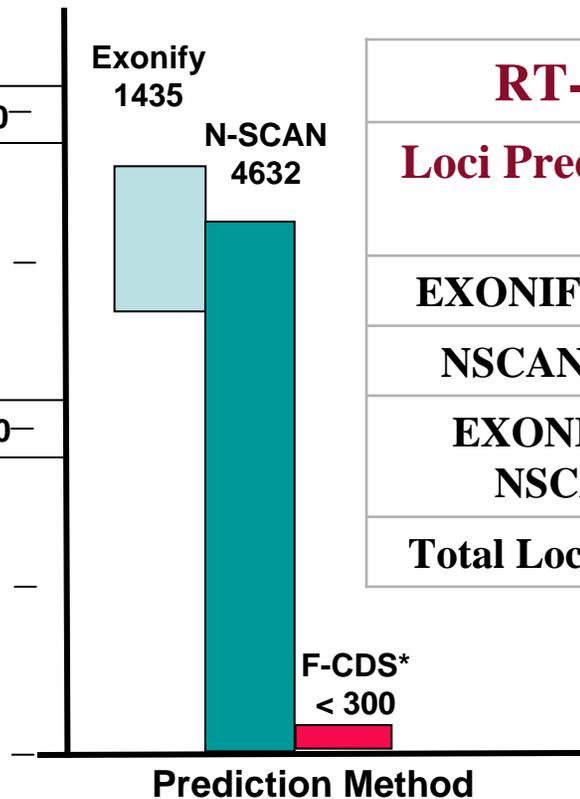
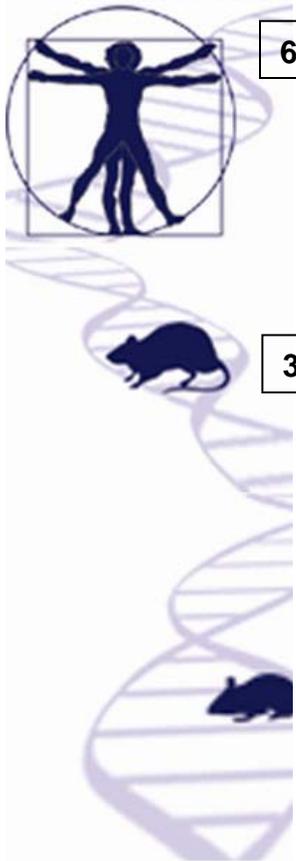


MGC: Phase III
Genes without (much)
experimental evidence

“C List” of Human Exon Predictions (M. Brent et al., & Haussler et. al.)

(as of January 2006)

Mammalian
Gene
Collection

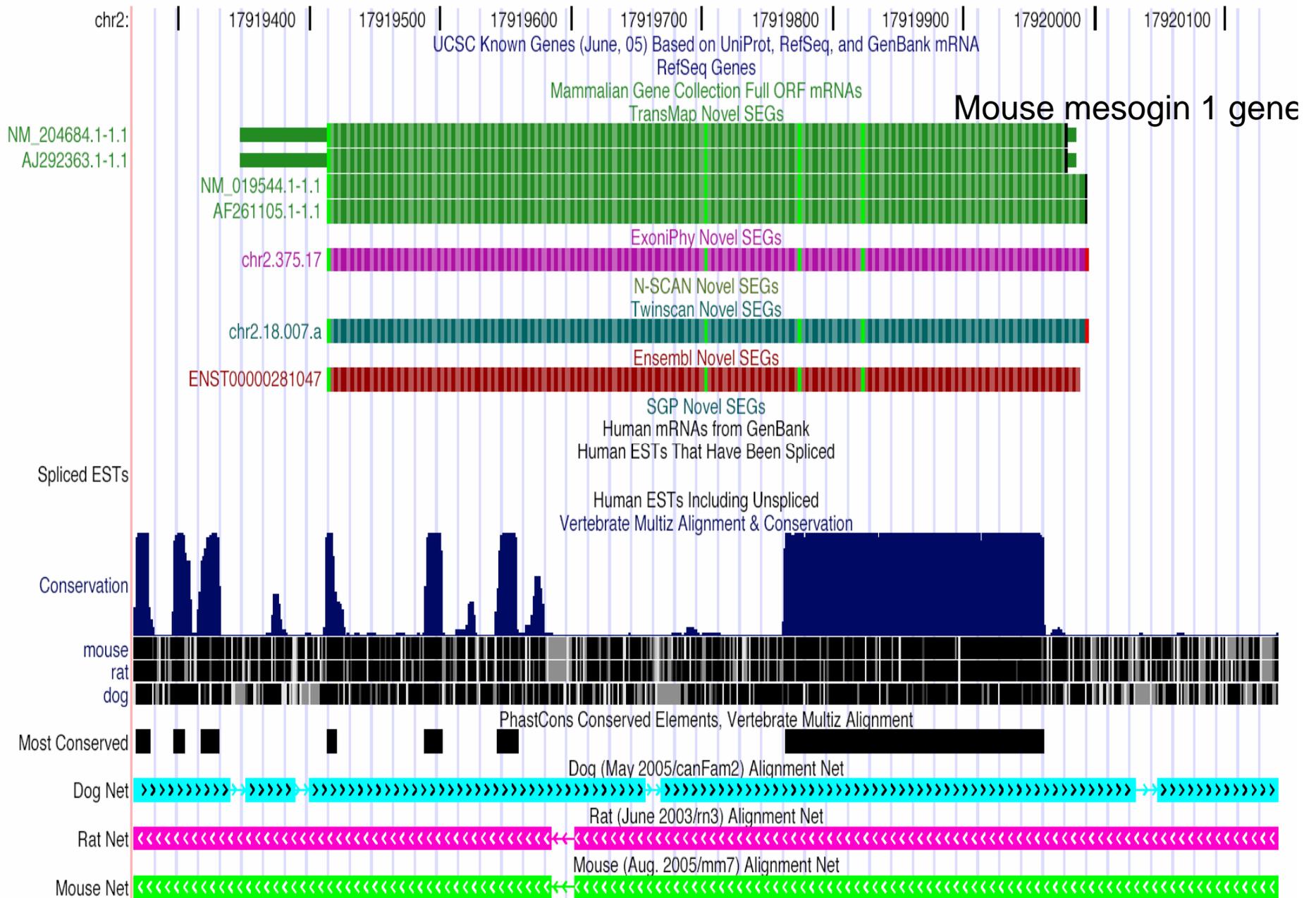


RT-PCR Verification of C List Predictions

Loci Predictions	No. Loci Tested	RT-PCR Success Rate
EXONIFY alone	16	19%
NSCAN alone	876	9%
EXONIFY + NSCAN	1219	41%
Total Loci Tested	2111	27%

About 500 predictions, which had at least parts of the transcript already verified by PCR were submitted to BCM for cloning

Novel Single Exon Gene Identification (D. Haussler & A. Siepel, UCSC & Cornell U.)



RT-PCR of the eight single exon genes (SEGs)

RT-PCR targets:

1. chr19.42.008
2. chr2.375.17
3. chr20.63.017
4. chr2.56.006
5. chr5.168.006
6. chr21.723.9
7. chr18.151.3
8. chr17.35.001

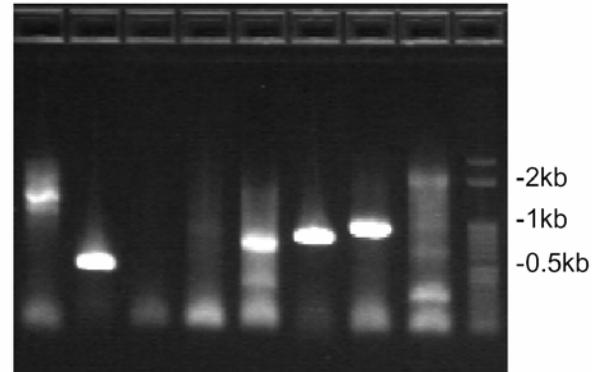
Expected sizes

(w/ gateway tails)

1. 404 bp
2. 470 bp
3. 614 bp
4. 638 bp
5. 776 bp
6. 812 bp
7. 962 bp
8. 1,197 bp

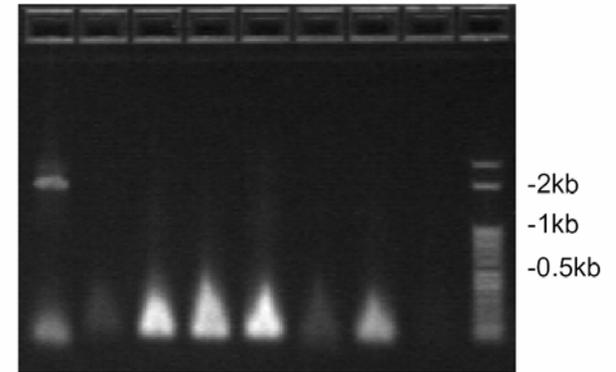
+RT (dT₁₆ priming)

1 2 3 4 5 6 7 8 M



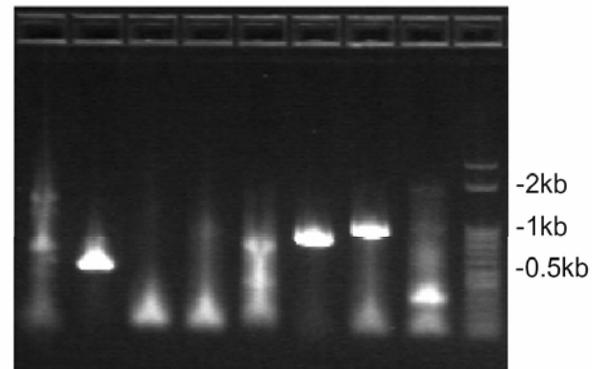
-RT control

1 2 3 4 5 6 7 8 M



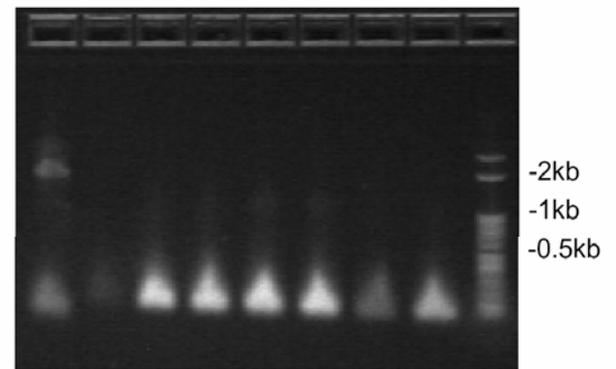
+RT (N₆ priming)

1 2 3 4 5 6 7 8 M

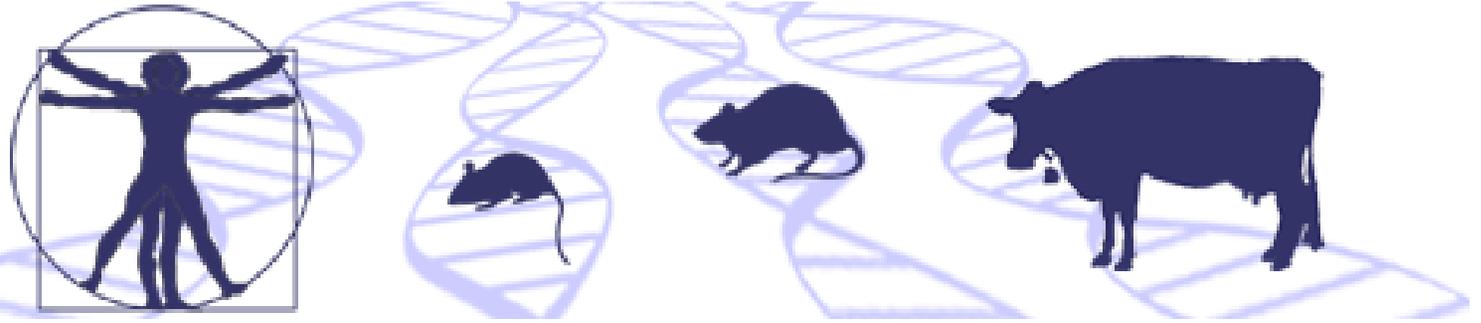


-RT control

1 2 3 4 5 6 7 8 M



Mammalian
Gene
Collection



MGC: Phase IV
In vitro Synthesis

DNA Synthesis Pilot

Mammalian
Gene
Collection



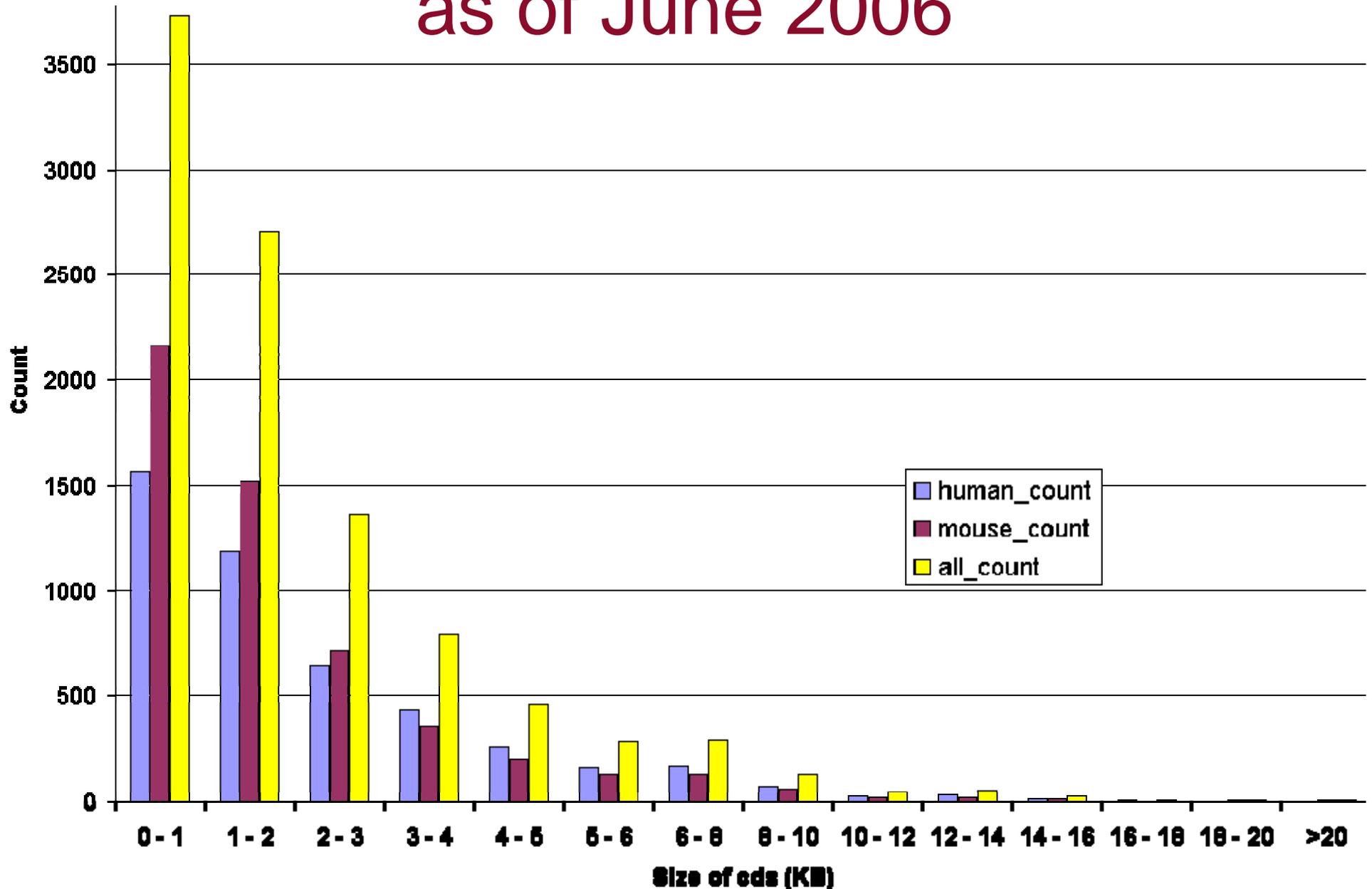
- 4 contractors

Each assigned a unique set of **18 targets** (0.5-6kb, total of 65 kb) + same 3 ORFs >6kb:

Factor V	NM_000130	cds= 6,675
APC	NM_000038	cds= 8,532
Dystrophin	NM_004006	cds=11,058

- pENTR223.1-Sfi or pDONR223.1
- All delivered sequence-verified clones, sequences, and sequence traces

Missing human and mouse genes as of June 2006



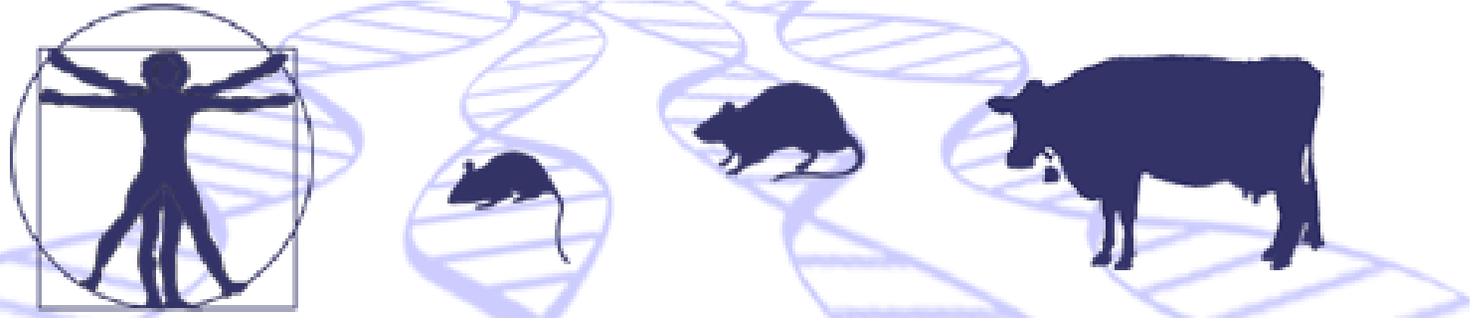
Completing the Human and Mouse Collection through Synthesis

Mammalian
Gene
Collection



- Estimated # of genes with RefSeq ID that will be verified by NCBI
 - ~2500 (almost equally divided between the 2 species)
- Expression ready vectors
- Exploring the option of 2 versions
 - with STOP codon
 - without STOP codon
- Size is directly proportional to cost

Mammalian
Gene
Collection



ORFeome Collection & MGC

ORFeome Collaboration

Mammalian
Gene
Collection



Current participants

- Dana Farber-CCSB (DF-CCSB)
- Deutsches Krebsforschungszentrum (DKFZ)
- Harvard Institute of Proteomics (HIP)
- IMAGE Consortium-LLNL
- Kazusa DNA Research Institute
- Mammalian Gene Collection (MGC)
- RIKEN Yokohama Institute
- Wellcome Trust-Sanger Institute (WTSI)

Human ORFeome Collaboration

Mammalian
Gene
Collection



Shared Goals:

- Complete collection of human full-ORF cDNA clones (for ~18,400 genes)*
- Expression-convenient format (Gateway Entry clones)
- ORFs + stop amino acids & - stop amino acid (for C-fusions)
 - Current plan – without STOP amino acid
- Sequence-verified clones
- Unrestricted worldwide availability

Mammalian
Gene
Collection



MGC: Its Status and Future

Interactome Networks

August 30, 2006

Daniela S. Gerhard, Ph.D.

Director, Office of Cancer Genomics,

National Cancer Institute, NIH